

Metabolic Gene Polymorphism Frequencies in Control Populations¹

Seymour Garte,² Laura Gaspari, Anna-Karin Alexandrie, Christine Ambrosone, Herman Autrup, Judith L. Autrup, Helena Baranova, Lise Bathum, Simone Benhamou, Paolo Boffetta, Christine Bouchardy, Katja Breskvar, Jurgen Brockmoller, Ingolf Cascorbi, Margie L. Clapper, Christiane Coutelle, Ann Daly, Marco Dell'Omo, Vita Dolzan, Carolyn M. Dresler, Anthony Fryer, Aage Haugen, David W. Hein, Allan Hildesheim, Ari Hirvonen, Ling-Ling Hsieh, Magnus Ingelman-Sundberg, Ivan Kalina, Daehee Kang, Masahiro Kihara, Chikako Kiyohara, Pierre Kremers, Philip Lazarus, Loïc Le Marchand, Maria Celeste Lechner, Ester M. M. van Lieshout, Stephanie London, Johannes J. Manni, Christine M. Maugard, Shunji Morita, Valle Nazar-Stewart, Kazumasa Noda, Yoshio Oda, Fritz F. Parl, Roberta Pastorelli, Irene Persson, Wilbert H. M. Peters, Agneta Rannug, Timothy Rebbeck, Angela Risch, Liliane Roelandt, Marjorie Romkes, David Ryberg, Jan Salagovic, Bernadette Schoket, Janeric Seidegard, Peter G. Shields, Edith Sim, Daniel Sinnet, Richard C. Strange, Isabel Stücker, Haruhiko Sugimura, Jordi To-Figueras, Paolo Vineis, Mimi C. Yu, and Emanuela Taioli

Environmental and Occupational Health Sciences Institute, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854 [S. G.]; Genetics Research Institute, Milan, Italy [S. G., L. G.]; Karolinska Institute, Stockholm, Sweden [A.-K. A., A. Ra., M. I.-S., I. P.]; National Toxicology Program, Jefferson, Arkansas [C. A.]; University of Aarhus, Aarhus, Denmark [H. A., J. L. A.]; Université d'Auvergne, Clermont-Ferrand, France [H. B.]; Odense University Hospital, Odense, Denmark [L. B.]; INSERM, Villejuif, France [S. B., C. B., I. S.]; IARC, Lyon, France [P. B.]; University of Ljubljana, Ljubljana, Slovenia [K. B., V. D.]; Universitatklínikum Charité-Humboldt, Universität zu Berlin, Berlin, Germany [J. B.]; Ernst Moritz Arndt University, Greifswald, Greifswald, Germany [I. C.]; Fox Chase Cancer Center, Philadelphia, Pennsylvania [M. L. C., C. M. D.]; Université de Bordeaux II, Bordeaux, France [C. C.]; University of Newcastle, Newcastle, United Kingdom [A. D.]; Istituto Medicina del Lavoro, Perugia, Italy [M. D.]; Keele University, Staffordshire, United Kingdom [A. F., R. C. S.]; National Institute of Occupational Health, Oslo, Norway [A. Ha., D. R.]; University of Louisville, Louisville, Kentucky [D. W. H.]; National Cancer Institute, Rockville, Maryland [A. Hil.]; Finnish Institute of Occupational Health, Helsinki, Finland [A. Hir.]; Chang Gung University, Tao-yuan, Taiwan [L.-L. H.]; P. J. Šafárik University, Košice, Slovakia [I. K., J. Sa.]; Seoul National University College of Medicine, Seoul, Korea [D. K.]; Kyoto University School of Public Health, Kyoto, Japan [M. K., K. N.]; Kyushu University, Fukuoka, Japan [C. K.]; Institut de Pathologie Liège-Belgium, Liege, Belgium [P. K., L. R.]; H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida [P. L.]; University of Hawaii-Honolulu, Honolulu, Hawaii [L. L. M.]; Universidade de Lisboa, Lisboa, Portugal [M. C. L.]; University Hospital, Nijmegen, the Netherlands [E. M. M. v. L., W. H. M. P.]; National Institute for Environmental Health Sciences, Research Triangle Park, North Carolina [S. L.]; University Hospital Maastricht, Maastricht, the

Netherlands [J. J. M.]; Centre René Gauducheau Nantes, Nantes-Saint Herblain, France [C. M. M.]; Yao Municipal Hospital, Osaka, Japan [S. M.]; Oregon Health Sciences University, Portland, Oregon [V. N.-S.]; Kanazawa University, Kanazawa-Ishikawa, Japan [Y. O.]; Vanderbilt University, Nashville, Tennessee [F. F. P.]; Mario Negri Institute, Milan, Italy [R. P.]; University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania [T. R.]; Deutsches Krebsforschungszentrum, Heidelberg, Germany [A. Ri.]; University of Pittsburgh, Pittsburgh, Pennsylvania [M. R.]; National Institute of Environmental Health, Budapest, Hungary [B. S.]; Lund University, Lund, Sweden [J. Se.]; Georgetown University Medical Center, Washington, D.C. [P. G. S.]; University of Oxford, Oxford, United Kingdom [E. S.]; Hôpital Sainte-Justine, Montreal, Quebec, Canada [D. S.]; Hamamatsu University School of Medicine, Hamamatsu, Shizuoka, Japan [H. S.]; Hospital Clinic Provincial, Barcelona, Spain [J. T.-F.]; Università degli Studi di Torino, Torino, Italy [P. V.]; University of Southern California/Norris Cancer Center, University of Southern California, Los Angeles, California [M. C. Y.]; and Ospedale Maggiore IRCCS-Direzione Scientifica, Milan, Italy [E. T.]

Abstract

Using the International Project on Genetic Susceptibility to Environmental Carcinogens (GSEC) database containing information on over 15,000 control (noncancer) subjects, the allele and genotype frequencies for many of the more commonly studied metabolic genes (*CYP1A1*, *CYP2E1*, *CYP2D6*, *GSTM1*, *GSTT1*, *NAT2*, *GSTP*, and *EPHX*) in the human population were determined. Major and significant differences in these frequencies were observed between Caucasians ($n = 12,525$), Asians ($n = 2,136$), and Africans and African Americans ($n = 996$), and some, but much less, heterogeneity was observed within Caucasian populations from different countries. No differences in allele frequencies were seen by age, sex, or type of controls (hospital patients versus population controls). No examples of linkage disequilibrium between the different loci were detected based on comparison of observed and expected frequencies for combinations of specific alleles.

Introduction

Increasing our understanding of the role of genetic factors in determining human susceptibility to the carcinogenic effects of environmental agents has become a major research goal in molecular epidemiology. The identification of high frequency (>1%) genetic polymorphisms in genes associated with carcinogen metabolism (1–4) has allowed the development of hypotheses that attempt to explain the high degree of individual variability in cancer susceptibility that has been observed (for example, among smokers).

Over recent years, numerous studies using case-control approaches and generally based on 100–300 cases have examined the association of one or a few polymorphisms with cancer risk (1, 5–8). Although progress has been made, many of these studies have produced conflicting results, in part because of the low penetrance of this category of susceptibility genes, resulting in insufficient power. Whereas the precise penetrance of these genes is not known (see other publications from the

Received 12/20/00; revised 6/8/01; accepted 6/25/01.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¹ Partially supported by European Commission Fund No. 96/CAN/33919.

² To whom requests for reprints should be addressed, at EOHHSI, 170 Frelinghuysen Road, Piscataway, NJ 08854.

GSEC³), it is clear that odds ratios of allelic variants rarely exceed 2–3 in the general population, which makes sample size a critical issue in case-control studies assessing the role of these genes in cancer.

To clarify the role of individual and composite genotypes at the most interesting and/or highly studied loci in cancer susceptibility, we began, in 1996, to gather data from investigators around the world on the frequencies of genetic polymorphisms of genes associated with carcinogen metabolism. A more detailed description of this project, the International Project on GSEC, has been recently published (9). It has been demonstrated in many studies that allele frequencies of the metabolic genes are not randomly distributed throughout the human population but follow diverse ethnic and/or geographic-specific patterns (10–13). However, no single published study to date on these gene polymorphisms has been large enough to precisely define the true population-specific frequency of most of these alleles in normal control populations. Furthermore, it has not yet been possible to determine whether other demographic variables are associated with specific allele frequencies. In this work, the contributed data from 52 laboratories representing 73 separate studies (both published and unpublished) have been pooled to characterize allele frequencies in 8 metabolic genes in a very large sample (15,843) of control subjects (defined as those individuals who served as the comparison group for subjects with cancer in case-control studies or groups of healthy individuals studied for other purposes) from different regions of the world.

Materials and Methods

Data were received from investigators who had been contacted or had learned of the study. Details of these aspects of the GSEC project, including response rate, contact strategy, the use of published and unpublished data, and definitions used for covariates were presented in a previous publication (9). Original data files were received for data that have been included in previous publications (14–67). None of the data included any personal identifiers. Noninformative consecutive identification numbers were assigned to each subject at the time of receipt of the data. It is therefore not possible to trace any particular subject in the database back to his/her actual identity through the identification number. All data on genotype were converted to a standard nomenclature as described in the accompanying letters by Garte *et al.* (68) and Ingelman Sundberg *et al.* (69), and in other reports (70–72). Data were received from the database in an Excel file, and all analyses were performed using SAS version 7.1 statistical software. For comparisons between groups, the χ^2 test was used. Multivariate analysis was used to assess the independent contribution of demographic factors such as race, age, sex, and national origin on allele frequency. Frequencies of gene polymorphisms were analyzed only when more than one study and more than 100 subjects were included for that polymorphism or allele.

The great majority of the data were generated by PCR analysis, although in a few older studies, some of the genotypes were determined by Southern blot analysis. In all cases, results were confirmed by PCR.

For CYP1A1, CYP2E1 and NAT2, all genotypes were in HWE with $P > 0.05$ by χ^2 test. Each individual study or data set was tested for HWE. Only one small data set received by the

Table 1 Information on metabolic gene polymorphisms in the database by ethnic group

Gene	Ethnicity	No. of studies	No. of subjects
CYP1A1	Caucasian	33	5434
	Asian	9	1144
	African	5	505
CYP2E1	Caucasian	16	1966
	Asian	5	719
	African	2	40
GSTM1	Caucasian	50	10514
	Asian	11	1511
	African	7	479
GSTT1	Caucasian	29	5577
	Asian	3	575
	African	3	89
GSTP1	Caucasian	13	2282
	Asian	1	243
	African	1	82
NAT2	Caucasian	21	3979
	Asian	1	36
	African	1	7
EH	Caucasian	5	922
	Asian	1	123
	African	1	21
CYP2D6	Caucasian	20	3530
	Asian	0	
	African	2	272

GSEC study was not included, because data were found not to be in HWE, and allele frequencies were far outside the expected ranges.

Results

The database of control subjects analyzed for metabolic gene polymorphisms consisted of 15,843 subjects. Of these, 12,525 were Caucasians (79.1%), 2,136 were Asians (13.5%), 936 were African Americans (5.9%), 60 were Africans (0.4%), 186 were of uncertain or other ethnicity (0.16%), and 75 had no ethnic information in the database. For all analyses, Africans and African Americans were combined, and those with uncertain or missing ethnic classification were excluded.

Table 1 shows the number of studies and the number of subjects according to their ethnicity for each gene contained in the database. Table 2 lists all of the data on the frequencies of the common polymorphisms for CYP1A1 and CYP2E1. Some of these data are from studies where only one polymorphism (for example only the MspI RFLP in CYP1A1) was examined. Because not all studies included analysis of both polymorphisms for each individual, not all of these data could be used for determining allele frequencies, which depend on information from two major polymorphic sites for each individual for CYP1A1 and CYP2E1. The data shown in the rest of the tables include only those individuals for whom polymorphism analysis was done on both sites. However, even in these cases, a true haplotype is not available because other less common polymorphic sites in these genes were usually not analyzed. For several genes (CYP1A1, CYP2E1, GSTM1, and GSTT1), there were sufficient data from more than one ethnic group to make comparisons between such groups as shown in Table 3. Frequencies of most of the alleles exhibited large differences between ethnic groups, especially for the CYP genes between Asians and Caucasians. For some of the genes, data was available only for Caucasians. For GSTP, EPHX, and CYP2D6, the

³ The abbreviations used are: GSEC, Genetic Susceptibility to Environmental Carcinogens; HWE, Hardy Weinberg equilibrium.

Table 2 Frequencies of *CYP1A1* and *CYP2E1* gene polymorphisms in different ethnic groups

Polymorphism	Caucasians No. (%)	Asians No. (%)	Africans No. (%)
<i>CYP1A1</i> MspI (<i>CYP1A1</i> *2A, <i>CYP1A1</i> *2B)			
No.	4453	638	461
Homozygous (wild-type)	3670 (82.4)	268 (42)	268 (58.1)
Heterozygous	729 (16.4)	281 (44)	166 (36)
Homozygous (variant)	54 (1.2)	89 (14)	27 (5.9)
<i>CYP1A1</i> exon 7 (<i>CYP1A1</i> *2B, <i>CYP1A1</i> *2C)			
No.	4790	1132	481
Homozygous (wild-type)	4319 (90.2)	670 (59.2)	456 (94.8)
Heterozygous	444 (9.3)	407 (36)	25 (5.2)
Homozygous (variant)	27 (0.6)	55 (4.9)	0 (0)
<i>CYP2E1</i> RsaI (<i>CYP2E1</i> *5A, <i>CYP2E1</i> *5B)			
No.	1454	719	NA ^a
Homozygous (wild-type)	1344 (92.4)	428 (59.5)	
Heterozygous	109 (7.5)	258 (35.9)	
Homozygous (variant)	1 (0.1)	33 (4.6)	
<i>CYP2E1</i> DraI (<i>CYP2E1</i> *5A, <i>CYP2E1</i> *6)			
No.	1360	286	NA
Homozygous (wild-type)	1162 (85.4)	138 (48.3)	
Heterozygous	187 (13.8)	121 (42.3)	
Homozygous (variant)	11 (0.8)	27 (9.4)	

^a NA, not available.

Table 3 Metabolic gene allele frequencies in different ethnic groups

Gene	Race	No. ^a	Heterozygous	Homozygous	Allele
<i>CYP1A1</i> *1	Caucasians	3814	0.190 (0.13–0.27) ^b	0.795 (0.71–0.87)	0.890
	Asians	626	0.460 (0.43–0.49)	0.395 (0.34–0.41)	0.625
	Africans	445	0.465	0.432	0.664
<i>CYP1A1</i> *2A	Caucasians	3814	0.105 (0.054–0.16)	0.005 (0–0.015)	0.058
	Asians	626	0.272 (0.23–0.31)	0.0128 (0–0.056)	0.149
	Africans	445	0.333	0.0517	0.218
<i>CYP1A1</i> *2B	Caucasians	3814	0.064 (0.025–0.12)	0.0001 (0–0.0057)	0.032
	Asians	626	0.331 (0.32–0.44)	0.0463 (0.031–0.058)	0.212
	Africans	445	0.036	0	0.018
<i>CYP1A1</i> *2C	Caucasians	3814	0.033 (0–0.095)	0.0021 (0–0.012)	0.0186
	Asians	626	0.027 (0–0.032)	0.0016 (0–0.012)	0.0152
	Africans	445	0.0135	0	0.00675
<i>CYP1A1</i> *3	Caucasians	735	0	0	0
	Africans	464	0.177	0.0043	0.0927
<i>CYP2E1</i> *5A	Caucasians	854	0.048 (0.034–0.095)	0.0012 (0–0.005)	0.0252
	Asians	286	0.367 (0.35–0.38)	0.0594 (0.054–0.63)	0.243
<i>CYP2E1</i> *5B	Caucasians	854	0.0105 (0–0.05)	0	0.00525
	Asians	286	0.021 (0.006–0.045)	0	0.0105
<i>CYP2E1</i> *6	Caucasians	854	0.102 (0.08–0.12)	0.0023 (0–0.0032)	0.0533
	Asians	286	0.126 (0.071–0.093)	0	0.0630
<i>EPHX</i> *3	Caucasians	685	0.398	0.117	0.316
<i>EPHX</i> *4	Caucasians	686	0.353	0.038	0.215
<i>GSTM1</i> *0	Caucasians	10514		0.531 (0.42–0.60)	
	Asians	1511		0.529 (0.42–0.54)	
	Africans	479		0.267 (0.16–0.36)	
<i>GSTT1</i> *0	Caucasians	5577		0.197 (0.13–0.26)	
	Asians	575		0.470 (0.35–0.52)	
<i>GSTM1</i> *0	Caucasians	5532		0.104	
<i>+T1</i> *0	Asians	407		0.246	
<i>GSTP1</i> *1	Caucasians	1137	0.493	0.438	0.685
<i>GSTP1</i> *2	Caucasians	1138	0.442	0.0413	0.262
<i>GSTP1</i> *3	Caucasians	878	0.126	0.0057	0.0687
<i>NAT2</i> *5	Caucasians	3847	0.482 (0.42–0.55)	0.219 (0.13–0.32)	0.46
<i>NAT2</i> *6	Caucasians	3618	0.430 (0.35–0.56)	0.070 (0.032–0.11)	0.285
<i>NAT2</i> *7	Caucasians	3129	0.055 (0.028–0.099)	0.0013 (0–0.071)	0.029

^a No. refers to the number of subjects tested.^b Numbers in parentheses give the range of values for individual studies used.

Table 4 Geographic distribution of four metabolic gene alleles in Caucasian populations

Country	<i>CYP1A1</i> *2A	<i>NAT2</i> *5	<i>GSTM1</i> *0	<i>GSTT1</i> *0
Canada	0.0602 (299) ^a	NA ^b	0.513 (304)	0.172 (274)
Denmark	NA	0.466 (426)	0.536 (537)	0.129 (358) ^c
Finland	0.0621 (145)	0.465 (414)	0.469 (482)	0.130 (385) ^c
France	0.0527 (171)	0.393 (244)	0.534 (1184)	0.168 (512)
Germany	0.0442 (882) ^c	0.461 (701)	0.516 (734)	0.195 (487)
Italy	0.0891 (303)	0.457 (550)	0.494 (810)	0.163 (553)
Netherlands	0.0335 (419) ^c	NA	0.504 (419)	0.229 (419)
Norway	0.0795 (107)	0.487 (371)	0.506 (423)	NA
Portugal	NA	0.457 (257)	0.583 (501)	NA
Saudi Arabia	NA	NA	0.563 (895)	NA
Slovakia	NA	NA	0.512 (332)	0.180 (322)
Slovenia	0.0654 (107)	NA	0.520 (102)	0.255 (102)
Spain	NA	NA	0.497 (312)	0.205 (312)
Sweden	0.0743 (512)	0.496 (420)	0.559 (544)	0.130 (423) ^c
UK	0.0694 (310)	NA	0.578 (1122) ^c	0.205 (922)
US	0.0662 (649)	0.432 (385)	0.543 (1751)	0.276 (286)

^a Numbers in parentheses denote the number of subjects tested.^b NA, not available; US, United States; UK, United Kingdom.^c Statistically significant difference ($P < 0.05$, by χ^2 test) compared with the rest of the population combined.

Table 5 Geographic distribution of metabolic gene alleles in Asian populations

	<i>GSTM1</i>	<i>GSTT1</i>	<i>CYP1A1</i> *2A
Japan	0.476 (639) ^{a,b}	0.353 (167) ^b	0.159 (330)
Korea	0.521 (165)	0.515 (165)	NA ^c
Singapore	0.562 (244)	0.519 (243)	NA

^a Numbers in parentheses denote the number of subjects tested.^b Statistically significant difference ($P < 0.05$, by χ^2 test) compared with the rest of the population combined.^c NA, not available.

database contains relatively sparse information at the present time, and frequencies for these genes are the most likely to be imprecise. This may be particularly true for *CYP2D6*, which is a complex gene with many alleles, only some of which have been included in the current database. More accurate estimates of *CYP2D6* allele frequencies may be found in other publications; therefore, this gene has not been included in Table 3. The range of values found in individual studies shown in Table 3 gives some idea of the degree of heterogeneity of the contributed data sets. As discussed in "Materials and Methods," the individual data sets were similar to the means presented here, and all were in HWE.

To determine whether regional gradients of allele frequency might exist within the Caucasian populations of Europe and North America, we analyzed allele frequencies for *GSTM1**0, *GSTT1**0, *CYP1A1**2A, and *NAT2**5 according to nation of origin, as shown in Table 4. For the *CYP1A1**2A allele, German and Dutch populations had a significantly lower frequency than the rest of Europe and North America, and for *GSTM1*, Great Britain had a higher frequency of deletion than the rest of the Caucasian population taken together. Although the frequency of *GSTM1* deletion in Portugal was also higher, this was not significant when compared to the rest of Europe and North America by χ^2 analysis. Three Scandinavian countries (Finland, Denmark, and Sweden) had a significant ($\chi^2 = 21.2$; $P < 0.001$) and substantially lower (30% less relative to the rest of Europe and North America) frequency of the *GSTT1* deletion.

Among Asians, a significant difference between Japanese

and other Asians was observed for both *GSTM1**0 and *GSTT1**0, with Japanese showing lower frequencies of both deletions (Table 5). However, the numbers of subjects for each of these countries was quite small, making any conclusions regarding heterogeneity within the Asian group premature.

There were 9399 (61.9%) men and 5790 (38.1%) women in the database. Asians had a higher proportion of men (77%) compared with Caucasians and Africans (59% and 57%, respectively). Information on sex was missing for 729 (4.6%) subjects. As shown in Table 6, no differences in allele frequencies by sex were seen among Caucasians for any gene except for *GSTT1*. Men appeared to have a lower deletion frequency than women. When the data were adjusted in a multivariate model by country of origin, there was no difference for *GSTT1* among Caucasians as a function of sex. Within the three Scandinavian countries, there was no difference in *GSTT1* deletion frequency with sex. For Asians, no differences with sex were observed for *GSTM1**0, *GSTT1**0, and *CYP1A1**2A. For *GSTT1*, the only data including women were from Singapore, therefore the data in Table 6 for *GSTT1**0 in Asians by sex refer to Singapore only.

In Africans, a significant difference was observed in the frequency of *GSTM1* homozygous deletions between men and women (Table 6). To better understand the source of the sex differences observed, multivariate analysis was used, and sex appeared to be the only significant variable for the observed difference in *GSTM1* polymorphism frequency.

The age distribution among the controls is exhibited in Fig. 1. Age data were missing for 1935 individuals. No differences were seen for any polymorphism in any of the genes for which sufficient information was available (*GSTM1*, *GSTT1*, and *CYP1A1*) in any racial group as a function of age. There appeared to be a trend toward higher rates of *GSTT1* deletion with increasing age (data not shown), but this was not significant.

We compared the frequencies of polymorphisms at the *CYP1A1*, *NAT2*, *GSTM1*, and *GSTT1* loci between controls drawn from hospital and population sources. Although there appeared to be differences according to source of controls for *GSTM1* in Caucasians and *GSTT1* in Asians (Table 7), after adjusting for sex and geographic area in a multivariate analysis,

Table 6 Metabolic gene allele frequencies as a function of sex

Gene	Ethnicity	Men (n)	Women (n)	Significance
<i>GSTM1</i> *0	Caucasian	0.526 (6015)	0.543 (4098)	NS ^a
	Asian	0.525 (1134)	0.541 (377)	NS
	African	0.233 (292)	0.321 (187)	<i>P</i> = 0.034
<i>GSTT1</i> *0	Caucasian	0.181 (3181)	0.210 (2074)	<i>P</i> = 0.010
	Non-Scandinavian	0.206 (2129)	0.215 (1960)	NS
	Scandinavian	0.131 (1052)	0.114 (114)	NS
	Asian (Malaysia)	0.482 (108)	0.548 (135)	NS
	African	NA ^b	NA	
<i>CYP1A1</i> *2A	Caucasian	0.0543 (2173)	0.0635 (1566)	NS
	Asian	0.148 (526)	0.150 (100)	NS
	African	0.218 (172)	0.225 (229)	NS

^a NS, no significant differences between men and women by χ^2 test.

^b NA, not available.

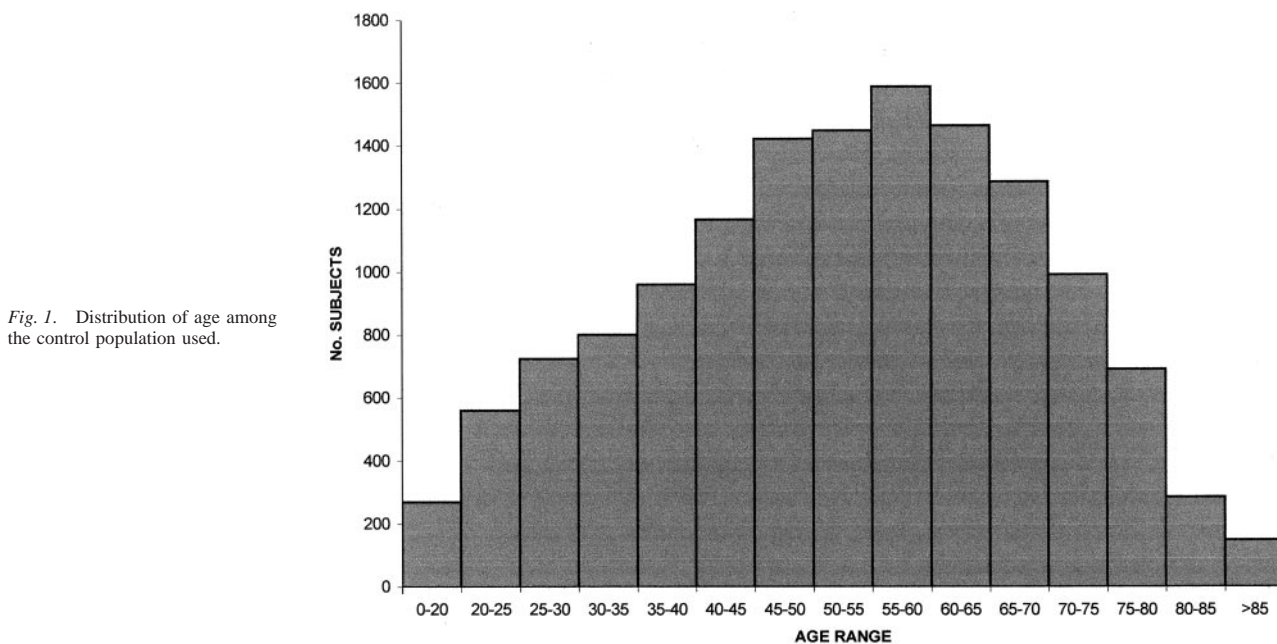


Fig. 1. Distribution of age among the control population used.

we found that there was no association between control type and genotype in either case.

Given the large number of subjects in the database, it was possible to estimate the genotype frequencies of certain metabolic genes. This was done for *CYP1A1* and *NAT2* for Caucasians (Table 8). Genotype assignments were made only for those subjects with available data on all the major polymorphisms. For the *NAT2* gene, genotypes were computed both using the data from almost 4000 subjects on the major allele groups (*5, 6, and 7) and from subtypes *5A, *5B, *5C, and *6A for a smaller set of subjects.

To test the hypothesis that there is no linkage disequilibrium between any of the loci examined here, we compared the observed frequencies of heterozygous and homozygous combinations of several of the alleles (*GSTM1**0, *GSTT1**0, *CYP1A1**2A, *CYP1A1**2B, *CYP1A1**5A, *CYP2E1**6, *NAT2**5, and *NAT2**6) with those expected from their population frequencies. All 25 possible double combinations for the different genes were examined separately for Caucasians and Asians. Some examples of these comparisons are shown in Table 9. In no case were any significant deviations from expected allele

combinations observed, suggesting that for these alleles, there is no linkage between any of the polymorphic alleles at these loci.

Discussion

Estimates of the population frequency of the polymorphic metabolic gene alleles have been reported in numerous publications. However, these estimates have almost always been from studies of a few hundred individuals at most. The expected imprecision of normal allele frequencies obtained from relatively few samples has sometimes led to erroneous conclusions regarding genetic heterogeneity according to geographic or other criteria. The availability of a large database on metabolic gene polymorphisms has allowed for a more precise estimate of the allele and genotype frequencies for many of these genes than has been previously possible. Such data could be of use to investigators for quality control purposes. For example, if in a smaller study (among Caucasians at least), control allele frequencies are observed to be significantly different from those

Table 7 Metabolic gene allele frequencies as a function of source of control population (healthy community or hospital)

Gene	Race	Community (n)	Hospital (n)	χ^2	Significance
<i>GSTM1</i> ^a	Caucasian	0.520 (5294)	0.546 (3206)	5.64	$P = 0.018$
<i>GSTM1</i>	African	0.279 (262)	0.254 (217)	0.38	NS ^b
<i>GSTM1</i>	Asian	0.522 (1265)	0.521 (165)	0.002	NS
<i>GSTT1</i>	Caucasian	0.183 (2735)	0.192 (2186)	0.64	NS
<i>GSTT1</i> ^a	Asian	0.519 (243)	0.434 (332)	4.05	$P = 0.04$
<i>CYP1A1</i> *2A	Caucasian	0.0598 (1964)	0.053 (1257)	0.45	NS
<i>CYP1A1</i> *2A	African	0.215 (386)	0.246 (59)	0.26	NS
<i>NAT2</i> *5	Caucasian	0.458 (2276)	0.468 (971)	0.22	NS
<i>CYP2E1</i> *5A	Caucasian	0.0153 (328)	0.0175 (343)	0.006	NS

^a No significant difference when national origin was included with control source in multivariate analysis.

^b NS, no significant difference between community and hospital controls.

Table 8 Genotype frequencies for *CYP1A1* and *NAT2*

A. <i>CYP1A1</i>	Caucasians (n = 3814)	Asians (n = 626)	Africans (n = 445)
*1/*1	0.795	0.395	0.432
*1/*2A	0.101	0.193	0.285
*1/*2C	0.0315	0.0208	0.0135
*1/*2B	0.0582	0.246	0.0315
*2A/*2A	0.0052	0.0128	0.0517
*2C/*2C	0.0021	0.0016	0
*2B/*2B	0.0013	0.0463	0
*2B/*2C	0.0016	0.0064	0
*2A/*2B	0.0045	0.0783	0.0045
*1/*3	0	0	0.135
*3/*3	0	0	0.0045
*2A/*3	0	0	0.0427
B. <i>NAT2</i>	Caucasians (n = 3846)	<i>NAT2</i> subtypes	Caucasians (n = 1164)
*4/*4	0.0725	*4/*5A	0.0306
*4/*5	0.210	*4/*5B	0.16
*4/*6	0.133	*4/*5C	0.017
*4/*7	0.0104	*5A/*5A	0.0596
*5/*5	0.219	*5B/*5B	0.117
*6/*6	0.0655	*5C/*5C	0.0036
*7/*7	0.001	*5A/*5B	0.024
*5/*6	0.255	*5A/*5C	0.0057
*5/*7	0.0174	*5B/*5C	0.0083
*6/*7	0.0169	*5A/*6A	0.0227
		*5B/*6A	0.216
		*5C/*6A	0.016

reported here, investigators might consider increasing their sample size or checking for methodological errors.

Sufficient data for several genes were available to estimate genotype frequencies in Caucasian and, in some cases, Asian populations. As was true for the allele frequencies, the genotype distribution for *CYP1A1* differed significantly between these two ethnic groups. In Asians and Africans, the so-called wild-type genotype *CYP1A1**1/*1 is in fact present in less than half of the population, which, along with similar situations for other genes such as *CYP2D6*, calls into question the very concept of the wild-type in human genetics, as has been discussed previously (73).

The frequency of the various *NAT2* genotypes has not been reported previously in any single study using populations of this size, so this represents the first useful analysis of these frequencies. It should be noted that not all alleles were evaluated in this analysis because sufficient data were lacking to make any improvement over existing published values for such alleles.

In addition to providing basic information on allele and genotype frequencies, we were also able to test certain hypotheses concerning genetic heterogeneity among and between populations. Because of previously observed differences in allele frequencies as a function of race, the population was divided into three groups, Caucasians, Asians, and Africans. The latter group included mostly African Americans as well as Africans. The group of Asians included a small number of Asian Americans. It should be stressed here that racial and ethnic identification is a difficult task, especially in situations where considerable admixture has been known to occur, and misclassification of individuals of mixed ancestry is very likely. Furthermore, defining ethnicity or race is probably not a biologically plausible way to divide the human population in terms of genetic differences (74). However, for the purposes of convenience and for hypothesis testing, we decided to perform the frequency analyses starting with conventional definitions of ethnicity. The often observed differences between population frequencies for the three major racial groupings were confirmed for most of the genes studied.

One of the hypotheses we examined was that the Caucasian population would be heterogeneous with respect to many of these alleles. We were able to test this hypothesis within the limits of the sample size for certain of the more commonly tested alleles. In general, there was very little heterogeneity among Caucasians, although we did find a small degree of heterogeneity between certain ethnic groups, with the largest and most significant example being the frequency of the *GSTT1* homozygous deletion in people of Scandinavian origin. It is not yet clear whether the differences in *CYP1A1**2A in the German and Dutch populations are due to true population differences or artifacts resulting from differences between laboratories. Furthermore, these differences were comparatively small and possibly not biologically meaningful; the same may be said for the difference in *GSTM1* deletion in the British and Portuguese. A more interesting difference, which probably reflects true population heterogeneity, was seen for *GSTT1* between Scandinavia and the rest of Europe and North America. This very clear 30% difference between Northern Europeans and Caucasians from the rest of Europe is not easily explained and may be important when comparing allele frequencies in case-control studies when subjects might be from different European origins. However, given the fact that allele frequencies did not vary very much among Caucasians, population stratification in studies of polymorphisms among European Americans is unlikely to be an important confounder (75).

Although a few differences in polymorphic allele frequencies were seen as a function of sex or source of the population (hospital *versus* population or other controls), most of these

Table 9 Examples of frequencies of observed genotype combinations compared to expected frequencies assuming nonlinkage

Genotype combination	Race	No.	Observed	Expected
<i>GSTM1</i> *0/*0 + <i>GSTT1</i> *0/*0	Caucasian	5532	0.104	0.105
<i>GSTM1</i> *0/*0 + <i>GSTT1</i> *0/*0	Asian	407	0.246	0.248
<i>GSTM1</i> *0/*0 + <i>CYP1A1</i> *1/*2A	Caucasian	3192	0.0573	0.0558
<i>GSTM1</i> *0/*0 + <i>CYP1A1</i> *1/*2A	Asian	509	0.132	0.144
<i>GSTM1</i> *0/*0 + <i>CYP1A1</i> *2A/*2A	Caucasian	3192	0.0025	0.0028
<i>GSTM1</i> *0/*0 + <i>CYP1A1</i> *1/*2B	Caucasian	3192	0.0326	0.0341
<i>GSTM1</i> *0/*0 + <i>CYP1A1</i> *1/*2B	Asian	509	0.165	0.175
<i>GSTM1</i> *0/*0 + <i>CYP1A1</i> *2B/*2B	Asian	509	0.0275	0.0245
<i>GSTM1</i> *0/*0 + <i>CYP2E1</i> *1/*5A	Asian	283	0.209	0.194
<i>GSTM1</i> *0/*0 + <i>NAT2</i> *5/*5	Caucasian	3266	0.122	0.116
<i>GSTM1</i> *0/*0 + <i>NAT2</i> *6/*6	Caucasian	3069	0.0401	0.0370
<i>GSTT1</i> *0/*0 + <i>CYP1A1</i> *1/*2A	Caucasian	2502	0.0164	0.0207
<i>GSTT1</i> *0/*0 + <i>CYP2E1</i> *1/*6	Caucasian	395	0.0253	0.0201
<i>CYP1A1</i> *1/*2A + <i>NAT2</i> *5/*5	Caucasian	1335	0.0217	0.0230
<i>CYP1A1</i> *1/*2B + (<i>NAT2</i> *4/*6 & <i>NAT2</i> *5/*6 & <i>NAT2</i> *6/*7) ^a	Caucasian	1151	0.0278	0.0276
<i>CYP2E1</i> *1/*6 + (<i>NAT2</i> *4/*5 & <i>NAT2</i> *5/*6 & <i>NAT2</i> *5/*7) ^b	Caucasian	409	0.0416	0.0491

^a Refers to all *NAT2**6 heterozygotes.^b Refers to all *NAT2**5 heterozygotes.

differences proved to be due to confounding factors such as geographic origin. An example was that of *GSTT1**0, which appeared to be higher in Caucasian women than men only because of the 10-fold prevalence of men in the Scandinavian studies. Multivariate analysis using the variables of sex, age, and country showed that the differences were only significant with respect to country and confirmed the lack of any effect of sex (as expected because none of these genes is located on the sex chromosomes) or source, with the exception of the difference in Africans for *GSTM1* as a function of sex. This difference was statistically significant but difficult to explain biologically. It must be noted that the population used for this analysis was relatively small (479 subjects); therefore, this result must be confirmed with a larger sample size before being accepted as resulting from some factor other than chance. No significant differences were seen with age, despite an intriguing trend for *GSTT1**0 to increase from childhood through maturity. We cannot speculate on the implications of a possible increase in *GSTT1* deletion with age; however, in a separate study, some of the authors have found a significant increase in *GSTT1* deletion among centenarians (76). The general lack of a significant association between allele frequency and age may allow epidemiologists to rule out the possibility that these polymorphisms are determinants of overall survival. The lack of any effect of choice of controls is important for comparison of different case-control studies that use one or the other source for the control population. These results also suggest that the use of hospital controls in studies of metabolic gene polymorphisms does not introduce bias related to genotype frequencies, although this work does not address other potential problems that may occur with hospital controls.

Because genetic susceptibility to environmental diseases probably must involve more than a single gene, it is useful to know whether any of these gene polymorphisms might be in linkage disequilibrium with each other. For example, if the *GSTM1* deletion and the *CYP1A1**2B allele were found together in the same subject at a frequency higher than expected from the independent frequencies of each polymorphism alone, it could be indicative of a linkage between these alleles. We observed no such evidence for linkage disequilibrium for any of the possible double combinations of the eight alleles examined

in noncancer controls. This is an important finding to serve as a control basis for analysis of such linkages in cancer cases. Of course, this is not a rigorous proof of the absence of linkage disequilibrium, and we have not tested for tri- or tetra-allelic disequilibrium. It should be noted that examples of possible linkage disequilibrium between certain metabolic genes were seen in earlier studies: *NAT1**10 was found to be associated with *NAT2**4 (77). There were not sufficient data on *NAT1* in the database to confirm this association.

In certain instances, the results presented here should be used with some caution. For example for *CYP1A1*, allele frequencies of the *2C allele are uncertain because there is still inconsistency in the results between different laboratories. In some laboratories, this allele is rarely or never seen, whereas in others, it is fairly common. This difference is almost definitely due to differences in laboratory methodology and should be resolved by interlaboratory exchanges of samples and methods. Although the population size used to make these estimates is larger than any previously used, for certain of the rarer alleles (such as *NAT2**7A), the paucity of the available data makes it difficult to estimate either allele or genotype frequencies.

It should also be emphasized that for most instances, the allele and genotype frequencies presented here do not always consider the complete spectrum of variants at a locus, due to limitations in the available data. Rare or newer alleles that have not yet been extensively analyzed (such as *CYP1A1**4, *NAT2**14, and so forth) were not considered. For *GSTM1**0 and *GSTT1**0, currently used methodology is unable to detect heterozygotes reliably, and therefore most studies did not present data on *GST* heterozygous deletions. This makes calculation of the *GST* deleted allele difficult. Newer methods (see below) will allow for detection of the heterozygous deletion. Furthermore, there are two *GSTM1* alleles, *GSTM1A* and *GSTM1B* (78), which future analyses will have to take into account.

Among other limitations of this study is the fact that information on race and age was collected in different ways by each of the investigators and is therefore not standardized. This is unlikely to have any effect on the results regarding age because very small errors are expected, and no associations were seen with age. As far as race is concerned, it is certainly possible that some misclassification occurred, given the diffi-

culty in making definitive assignments on race as discussed above. However, all cases where race was either unknown or unclear were excluded from the analysis.

Publication bias is always a possible limitation of combining data from various sources as in a typical meta-analysis. This may be less problematic in our work because unpublished data sets were also requested and included in the total data set.

Differences in laboratory techniques for analysis of genotype are probably not a major source of error, because most of the PCR-based techniques currently used for such assignments have become standardized. One exception noted above is the use of allele-specific *versus* restriction site PCR for detection of the *CYP1A1* mutation 2455A>G in exon 7, which is contained in *CYP1A1* alleles *2B and *2C. Furthermore, for *NAT2*, many of the commonly used PCR techniques do not assess all of the identified polymorphisms. Newer high-throughput techniques using fluorescent technologies or microarrays (79) will have the capacity to produce data on genotype much more efficiently than has been done to date, but standardization and common usage of these new methods have yet to be achieved. The analysis of polymorphisms in drug-metabolizing genes will have an important role in establishing a panel of single nucleotide polymorphisms that have known functional significance in post-genome analysis, not only in determining the role of xenobiotics in cancer, but also in other multifactorial disorders where environmental factors may be involved.

Because the information in the GSEC database continues to grow as more investigators become participants in the study, and new genes are added, it should be possible in the near future to update the results presented here and to be more certain of the true population frequencies. It is especially desirable that more data will be forthcoming from Asian and African populations because these have been relatively underrepresented thus far compared with data on Caucasians. With new methods of high-throughput analysis, DNA samples from very large cohorts (>100,000) may be used for detection of multiple allele frequencies in a very efficient manner. It will be interesting to compare frequencies determined by these methods with those reported here, which were generally determined using more standard PCR methods. Until such catalogues of allele frequencies from hundreds of thousands of subjects are available, this report presents the largest and most accurate estimate to date of these frequencies in healthy populations.

Acknowledgments

We thank Dr. Daniella Marinelli for collaboration with the GSEC Project.

References

- Vineis, P., Malats, N., Lang, M., D'Errico, A., Caporaso, N., Cuzick, J., and Boffetta, P. (eds.) *Metabolic Polymorphisms and Susceptibility to Cancer*. IARC Publications, Lyon, France, 1999.
- Hein, D. W., Doll, M. A., Fretland, A. J., Leff, M. A., Webb, S. J., Xiao, G. H., Devanaboyina, U. S., Nangju, N. A., and Feng, Y. Molecular genetics and epidemiology of the *NAT1* and *NAT2* acetylation polymorphisms. *Cancer Epidemiol. Biomark. Prev.*, 9: 29–42, 2000.
- Cascorbi, I., Brockmoller, J., Mrozkiewicz, P. M., Muller, A., and Roots, I. Arylamine *N*-acetyltransferase activity in man. *Drug Metab. Rev.*, 31: 489–502, 1999.
- Clapper, M. L., and Szarka, C. E. Glutathione *S*-transferases—biomarkers of cancer risk and chemopreventive response. *Chemico-Biol. Interactions*, 111–112: 377–388, 1998.
- Strange, R. C., Lear, J. T., and Fryer, A. A. Polymorphism in glutathione *S*-transferase loci as a risk factor for common cancers. *Archiv. Toxicol. Suppl.*, 20: 419–428, 1998.
- Hirvonen, A. Polymorphisms of xenobiotic-metabolizing enzymes and susceptibility to cancer. *Environ. Health Perspect.*, 107 (Suppl. 1): 37–47, 1999.

- Autrup, H. Genetic polymorphisms in human xenobiotica metabolizing enzymes as susceptibility factors in toxic response. *Mutat. Res.*, 464: 65–76, 2000.
- Bartsch, H., Nair, U., Risch, A., Rojas, M., Wikman, H., and Alexandrov, K. Genetic polymorphism of CYP genes, alone or in combination, as a risk modifier of tobacco-related cancers. *Cancer Epidemiol. Biomark. Prev.*, 9: 3–28, 2000.
- Taioli, E. International collaborative study on genetic susceptibility to environmental carcinogens. *Cancer Epidemiol. Biomark. Prev.*, 8: 727–728, 1999.
- Garte, S. The role of ethnicity in cancer susceptibility gene polymorphisms: the example of *CYP1A1*. *Carcinogenesis (Lond.)*, 19: 1329–1332, 1998.
- London, S. J., Smart, J., and Daly, A. K. Lung cancer risk in relation to genetic polymorphisms of microsomal epoxide hydrolase among African-Americans and Caucasians in Los Angeles County. *Lung Cancer*, 28: 147–155, 2000.
- Watson, M. A., Stewart, R. K., Smith, G. B., Massey, T. E., and Bell, D. A. Human glutathione *S*-transferase P1 polymorphisms: relationship to lung tissue enzyme activity and population frequency distribution. *Carcinogenesis (Lond.)*, 19: 275–280, 1998.
- Weber, W. W. Populations and genetic polymorphisms. *Molec. Diagnosis*, 4: 299–307, 1999.
- Ambrosone, C. B., Coles, B. F., Freudenheim, J. L., and Shields, P. G. Glutathione-*S*-transferase (GSTM1) genetic polymorphisms do not affect human breast cancer risk, regardless of dietary antioxidants. *J. Nutr.*, 129 (2S Suppl.): 565S–568S, 1999.
- Ambrosone, C. B., Freudenheim, J. L., Sinha, R., Graham, S., Marshall, J. R., Vena, J. E., Laughlin, R., Nemoto, T., and Shields, P. G. Breast cancer risk, meat consumption and *N*-acetyltransferase (*NAT2*) genetic polymorphisms. *Int. J. Cancer*, 75: 825–830, 1998.
- Saarikoski, S. T., Sata, F., Husgafvel-Pursiainen, K., Rautalahti, M., Haukka, J., Impivaara, O., Jarvisalo, J., Vainio, H., and Hirvonen, A. CYP2D6 ultrarapid metabolizer genotype as a potential modifier of smoking behaviour. *Pharmacogenetics*, 10: 5–10, 2000.
- Matthias, C., Bockmuhl, U., Jahnke, V., Jones, P. W., Hayes, J. D., Alldersea, J., Gilford, J., Bailey, L., Bath, J., Worrall, S. F., Hand, P., Fryer, A. A., and Strange, R. C. Polymorphism in cytochrome P450 CYP2D6, CYP1A1, CYP2E1 and glutathione *S*-transferase, GSTM1, GSTM3, GSTT1 and susceptibility to tobacco-related cancers: studies in upper aerodigestive tract cancers. *Pharmacogenetics*, 8: 91–100, 1998.
- Leathart, J. B., London, S. J., Steward, A., Adams, J. D., Idle, J. R., and Daly, A. K. CYP2D6 phenotype-genotype relationships in African-Americans and Caucasians in Los Angeles. *Pharmacogenetics*, 8: 529–541, 1998.
- Hsieh, L. L., Huang, R. C., Yu, M. W., Chen, C. J., and Liaw, Y. F. L-myc, GST M1 genetic polymorphism and hepatocellular carcinoma risk among chronic hepatitis B carriers. *Cancer Lett.*, 103: 171–176, 1996.
- Jacquet, M., Lambert, V., Baudoux, E., Muller, M., Kremers, P., and Gielen, J. Correlation between P450 CYP1A1 inducibility, *MspI* genotype and lung cancer incidence. *Eur. J. Cancer*, 32A: 1701–1706, 1996.
- Nazar-Stewart, V., Motulsky, A. G., Eaton, D. L., White, E., Hornung, S. K., Leng, Z. T., Stapleton, P., and Weiss, N. S. The glutathione *S*-transferase μ polymorphism as a marker for susceptibility to lung carcinoma. *Cancer Res.*, 53: 2313–2318, 1993.
- Warholm, M., Rane, A., Alexandrie, A. K., Monaghan, G., and Rannug, A. Genotypic and phenotypic determination of polymorphic glutathione transferase T1 in a Swedish population. *Pharmacogenetics*, 5: 252–254, 1995.
- Rebeck, T. R., Rosvold, E. A., Duggan, D. J., Zhang, J., and Buetow, K. H. Genetics of CYP1A1: coamplification of specific alleles by polymerase chain reaction and association with breast cancer. *Cancer Epidemiol. Biomark. Prev.*, 3: 511–514, 1994.
- Risch, A., Wallace, D. M., Bathers, S., and Sim, E. Slow *N*-acetylation genotype is a susceptibility factor in occupational and smoking related bladder cancer. *Hum. Mol. Genet.*, 4: 231–236, 1995.
- Stucker, I., de Waziers, I., Cenee, S., Bignon, J., Depierre, A., Milleron, B., Beaune, P., and Hemon, D. GSTM1, smoking and lung cancer: a case-control study. *Int. J. Epidemiol.*, 28: 829–835, 1999.
- Legrand-Andreoletti, M., Stucker, I., Marex, D., Galais, P., Cosme, J., Sabbagh, N., Spire, C., Cenee, S., Lafitte, J. J., Beaune, P., and Broly, F. Cytochrome P450 CYP2D6 gene polymorphism and lung cancer susceptibility in Caucasians. *Pharmacogenetics*, 8: 7–14, 1998.
- Le Marchand, L., Sivaraman, L., Pierce, L., Seifried, A., Lum, A., Wilkens, L. R., and Lau, A. F. Associations of CYP1A1, GSTM1, and CYP2E1 polymorphisms with lung cancer suggest cell type specificities to tobacco carcinogens. *Cancer Res.*, 58: 4858–4863, 1998.
- Wadelius, M., Autrup, J. L., Stubbins, M. J., Andersson, S. O., Johansson, J. E., Wadelius, C., Wolf, C. R., Autrup, H., and Rane, A. Polymorphisms in *NAT2*, *CYP2D6*, *CYP2C19* and *GSTP1* and their association with prostate cancer. *Pharmacogenetics*, 9: 333–340, 1999.

29. Autrup, J. L., Thomassen, L. H., Olsen, J. H., Wolf, H., and Autrup, H. Glutathione *S*-transferases as risk factors in prostate cancer. *Eur. J. Cancer Prev.*, 8: 525–532, 1999.
30. Rodriguez, J. W., Kirilin, W. G., Ferguson, R. J., Doll, M. A., Gray, K., Rustan, T. D., Lee, M. E., Kemp, K., Urso, P., and Hein, D. W. Human acetyltransferase genotype: relationship to colorectal cancer incidence and arylamine *N*-acetyltransferase expression in colon cytosol. *Arch. Toxicol.*, 67: 445–452, 1993.
31. To-Figueras, J., Gene, M., Gomez-Catalan, J., Galan, M. C., Fuentes, M., Ramon, J. M., Rodamilans, M., Huguet, E., and Corbella, J. Glutathione *S*-transferase M1 (GSTM1) and T1 (GSTT1) polymorphisms and lung cancer risk among Northwestern Mediterraneans. *Carcinogenesis (Lond.)*, 18: 1529–1533, 1997.
32. Peluso, M., Airolidi, L., Armelle, M., Martone, T., Coda, R., Malaveille, C., Giacomelli, G., Terrone, C., Casetta, G., and Vineis, P. White blood cell DNA adducts, smoking, and NAT2 and GSTM1 genotypes in bladder cancer: a case-control study. *Cancer Epidemiol. Biomark. Prev.*, 7: 341–346, 1998.
33. Baranova, H., Perriot, J., Albuissou, E., Ivaschenko, T., Baranov, V. S., Hemery, B., Mouraire, P., Riou, N., and Malet, P. Peculiarities of the GSTM1 0/0 genotype in French heavy smokers with various types of chronic bronchitis. *Hum. Genet.*, 99: 822–826, 1997.
34. Jourenkova-Mironova, N., Wikman, H., Bouchardy, C., Voho, A., Dayer, P., Benhamou, S., and Hirvonen, A. Role of glutathione *S*-transferase GSTM1, GSTM3, GSTP1 and GSTT1 genotypes in modulating susceptibility to smoking-related lung cancer. *Pharmacogenetics*, 8: 495–502, 1998.
35. Ruas, J. L., and Lechner, M. C. Allele frequency of CYP2C19 in a Portuguese population. *Pharmacogenetics*, 7: 333–335, 1997.
36. Park, L. Y., Muscat, J. E., Kaur, T., Schantz, S. P., Stern, J. C., Richie, J. P., Jr., and Lazarus, P. Comparison of GSTM polymorphisms and risk for oral cancer between African-Americans and Caucasians. *Pharmacogenetics*, 10: 123–131, 2000.
37. Krajinovic, M., Labuda, D., Richer, C., Karimi, S., and Sinnett, D. Susceptibility to childhood acute lymphoblastic leukemia: influence of CYP1A1, CYP2D6, GSTM1, and GSTT1 genetic polymorphisms. *Blood*, 93: 1496–1501, 1999.
38. Bathum, L., Andersen-Ranberg, K., Boldsen, J., Brosen, K., and Jeune, B. Genotypes for the cytochrome P450 enzymes CYP2D6 and CYP2C19 in human longevity. Role of CYP2D6 and CYP2C19 in longevity. *Eur. J. Clin. Pharmacol.*, 54: 427–430, 1998.
39. Seow, A., Shi, C. Y., Chung, F. L., Jiao, D., Hankin, J. H., Lee, H. P., Coetzee, G. A., and Yu, M. C. Urinary total isothiocyanate (ITC) in a population-based sample of middle-aged and older Chinese in Singapore: relationship with dietary total ITC and glutathione *S*-transferase M1/T1/P1 genotypes. *Cancer Epidemiol. Biomark. Prev.*, 7: 775–781, 1998.
40. Charrier, J., Maugard, C. M., Le Mevel, B., and Bignon, Y. J. Allelotype influence at glutathione *S*-transferase M1 locus on breast cancer susceptibility. *Br. J. Cancer*, 79: 346–353, 1999.
41. Kiyohara, C., Hirohata, T., and Inutsuka, S. The relationship between aryl hydrocarbon hydroxylase and polymorphisms of the *CYP1A1* gene. *Jpn. J. Cancer Res.*, 87: 18–24, 1996.
42. Bailey, L. R., Roodi, N., Verrier, C. S., Yee, C. J., Dupont, W. D., and Parl, F. F. Breast cancer and CYP1A1, GSTM1, and GSTT1 polymorphisms: evidence of a lack of association in Caucasians and African Americans. *Cancer Res.*, 58: 65–70, 1998.
43. Salagovic, J., Kalina, I., Stubna, J., Habalova, V., Hrivnak, M., Valansky, L., Kohut, A., and Biros, E. Genetic polymorphism of glutathione *S*-transferases M1 and T1 as a risk factor in lung and bladder cancers. *Neoplasma*, 45: 312–317, 1998.
44. Hildesheim, A., Anderson, L. M., Chen, C. J., Cheng, Y. J., Brinton, L. A., Daly, A. K., Reed, C. D., Chen, I. H., Caporaso, N. E., Hsu, M. M., Chen, J. Y., Idle, J. R., Hoover, R. N., Yang, C. S., and Chhabra, S. K. CYP2E1 genetic polymorphisms and risk of nasopharyngeal carcinoma in Taiwan. *J. Natl. Cancer Inst. (Bethesda)*, 89: 1207–1212, 1997.
45. Coutelle, C., Ward, P. J., Fleury, B., Quattrocchi, P., Chambrin, H., Iron, A., Couzigou, P., and Cassaigne, A. Laryngeal and oropharyngeal cancer, and alcohol dehydrogenase 3 and glutathione *S*-transferase M1 polymorphisms. *Hum. Genet.*, 99: 319–325, 1997.
46. Romkes, M., Chern, H. D., Lesnick, T. G., Becich, M. J., Persad, R., Smith, P., and Branch, R. A. Association of low CYP3A activity with p53 mutation and CYP2D6 activity with Rb mutation in human bladder cancer. *Carcinogenesis (Lond.)*, 17: 1057–1062, 1996.
47. Brockmoller, J., Cascorbi, I., Kerb, R., and Roots, I. Combined analysis of inherited polymorphisms in arylamine *N*-acetyltransferase 2, glutathione *S*-transferases M1 and T1, microsomal epoxide hydrolase, and cytochrome P450 enzymes as modulators of bladder cancer risk. *Cancer Res.*, 56: 3915–3925, 1996.
48. Dolzan, V., Rudolf, Z., and Breskvar, K. Human CYP2D6 gene polymorphism in Slovene cancer patients and healthy controls. *Carcinogenesis*, 16: 2675–2678, 1995.
49. Katoh, T., Inatomi, H., Nagaoka, A., and Sugita, A. Cytochrome P4501A1 gene polymorphism and homozygous deletion of the glutathione *S*-transferase M1 gene in urothelial cancer patients. *Carcinogenesis (Lond.)*, 16: 655–657, 1995.
50. Kihara, M., Kihara, M., and Noda, K. Risk of smoking for squamous and small cell carcinomas of the lung modulated by combinations of CYP1A1 and GSTM1 gene polymorphisms in a Japanese population. *Carcinogenesis (Lond.)*, 16: 2331–2336, 1995.
51. Oda, Y., Tanaka, M., and Nakanishi, I. Relation between the occurrence of *K-ras* gene point mutations and genotypes of polymorphic *N*-acetyltransferase in human colorectal carcinomas. *Carcinogenesis (Lond.)*, 15: 1365–1369, 1994.
52. Shields, P. G., Caporaso, N. E., Falk, R. T., Sugimura, H., Trivers, G. E., Trump, B. F., Hoover, R. N., Weston, A., and Harris, C. C. Lung cancer, race, and a CYP1A1 genetic polymorphism. *Cancer Epidemiol. Biomark. Prev.*, 2: 481–485, 1993.
53. Sugimura, H., Wakai, K., Genka, K., Nagura, K., Igarashi, H., Nagayama, K., Ohkawa, A., Baba, S., Morris, B. J., Tsygane, S., Ohno, Y., Gao, C., Li, Z., Takezaki, T., Tajima, K., and Iwamasa, T. Association of Ile462Val (exon 7) polymorphism of cytochrome P450 1A1 with lung cancer in the Asian population: further evidence from a case-control study in Okinawa. *Cancer Epidemiol. Biomark. Prev.*, 7: 413–417, 1998.
54. Evans, D. A., Seidegard, J., and Narayanan, N. The GSTM1 genetic polymorphism in healthy Saudi Arabians and Filipinos, and Saudi Arabians with coronary atherosclerosis. *Pharmacogenetics*, 6: 365–367, 1996.
55. Cascorbi, I., Brockmoller, J., and Roots, I. A C4887A polymorphism in exon 7 of human CYP1A1: population frequency, mutation linkages, and impact on lung cancer susceptibility. *Cancer Res.*, 56: 4965–4969, 1996.
56. Henning, S., Cascorbi, I., Munchow, B., Jahnke, V., and Roots, I. Association of arylamine *N*-acetyltransferases NAT1 and NAT2 genotypes to laryngeal cancer risk. *Pharmacogenetics*, 9: 103–111, 1999.
57. Malats, N., Camus-Radon, A. M., Nyberg, F., Ahrens, W., Constantinescu, V., Mukeria, A., Benhamou, S., Batura-Gabryel, H., Bruske-Hohlfeld, I., Simonato, L., Menezes, A., Lea, S., Lang, M., and Boffetta, P. Lung cancer risk in nonsmokers and GSTM1 and GSTT1 genetic polymorphism. *Cancer Epidemiol. Biomark. Prev.*, 9: 827–833, 2000.
58. Tefre, T., Ryberg, D., Haugen, A., Nebert, D. W., Skaug, V., Brogger, A., and Borresen, A. L. Human CYP1A1 (cytochrome P(1)450) gene: lack of association between the Msp I restriction fragment length polymorphism and incidence of lung cancer in a Norwegian population. *Pharmacogenetics*, 1: 20–25, 1991.
59. Ryberg, D., Skaug, V., Hewer, A., Phillips, D. H., Harries, L. W., Wolf, C. R., Ogreid, D., Ulvik, A., Vu, P., and Haugen, A. Genotypes of glutathione transferase M1 and P1 and their significance for lung DNA adduct levels and cancer risk. *Carcinogenesis (Lond.)*, 18: 1285–1289, 1997.
60. van Lieshout, E. M., Roelofs, H. M., Dekker, S., Mulder, C. J., Wobbes, T., Jansen, J. B., and Peters, W. H. Polymorphic expression of the glutathione *S*-transferase P1 gene and its susceptibility to Barrett's esophagus and esophageal carcinoma. *Cancer Res.*, 59: 586–589, 1999.
61. Schoket, B., Phillips, D. H., Kostic, S., and Vincze, I. Smoking-associated bulky DNA adducts in bronchial tissue related to CYP1A1 MspI and GSTM1 genotypes in lung patients. *Carcinogenesis (Lond.)*, 19: 841–846, 1998.
62. Taioli, E., Crofts, F., Demopoulos, R., Trachman, J., Toniolo, P., and Garte, S. J. A specific African American CYP1A1 polymorphism is associated with adenocarcinoma of the lung. *Cancer Res.*, 55: 472–473, 1995.
63. Taioli, E., Crofts, F., Trachman, J., Demopoulos, R., Ford, J., and Garte, S. J. Enhanced risk of lung and prostate adenocarcinoma in African-American smokers carrying a novel cyp1A1 gene polymorphism. *Poly. Arom. Comp.*, 11: 19–26, 1996.
64. Taioli, E., Ford, J., Trachman, J., Li, Y., Demopoulos, R., and Garte, S. Lung cancer risk and CYP1A1 genotype in African Americans. *Carcinogenesis (Lond.)*, 19: 813–817, 1998.
65. Persson, I., Johansson, I., Lou, Y. C., Yue, Q. Y., Duan, L. S., Bertilsson, L., and Ingelman-Sundberg, M. Genetic polymorphism of xenobiotic metabolizing enzymes among Chinese lung cancer patients. *Int. J. Cancer*, 81: 325–329, 1999.
66. Morita, S., Yano, M., Shiozaki, H., Tsujinaka, T., Ebisui, C., Morimoto, T., Kishibuti, M., Fujita, J., Ogawa, A., Taniguchi, M., Inoue, M., Tamura, S., Yamazaki, K., Kikkawa, N., Mizunoya, S., and Monden, M. CYP1A1, CYP2E1 and GSTM1 polymorphisms are not associated with susceptibility to squamous-cell carcinoma of the esophagus. *Int. J. Cancer*, 71: 192–195, 1997.
67. Oude Ophuis, M. B., van Lieshout, E. M., Roelofs, H. M., Peters, W. H., and Manni, J. J. Glutathione *S*-transferase M1 and T1 and cytochrome P4501A1 polymorphisms in relation to the risk for benign and malignant head and neck lesions. *Cancer (Phila.)*, 82: 936–943, 1998.
68. Garte, S., Boffetta, P., Caporaso, N., and Vineis, P. Metabolic gene allele nomenclature. *Cancer Epidemiol. Biomark. Prev.*, 10: 1305–1306, 2001.

69. Ingelman-Sundberg, M., Oscarson, M., Daly, A. K., Garte, S., and Nebert, D. W. Human cytochrome P-450 (CYP) genes: a web page for the nomenclature of alleles. *Cancer Epidemiol. Biomark. Prev.*, 10: 1307–1308, 2001.
70. Vatsis, K. P., Weber, W. W., Bell, D. A., Dupret, J.-M., Price-Evans, D. A., Grant, D. M., Hein, D. W., Lin, H. J., Meyer, U. A., Relling, M. V., Sim, E., Suzuki, T., and Yamazoe, Y. Nomenclature for *N*-acetyltransferases. *Pharmacogenetics*, 5: 1–17, 1995.
71. Hein, D. W., Grant, D. M., and Sim, E. Update on consensus arylamine *N*-acetyltransferase gene nomenclature. *Pharmacogenetics*, 10: 291–292, 2000.
72. Ingelman-Sundberg, M., Daly, A. K., Oscarson, M., and Nebert, D. W. Human cytochrome P450 (CYP) genes: recommendations for the nomenclature of alleles. *Pharmacogenetics*, 10: 91–93, 2000.
73. Nebert, D. W. Suggestions for the nomenclature of human alleles: relevance to ecogenetics, pharmacogenetics and molecular epidemiology. *Pharmacogenetics*, 10: 279–290, 2000.
74. Cavalli-Sforza, L. L. *Genes, Peoples and Languages*. New York: Farrar Straus & Giroux, 2000.
75. Wacholder, S., Rothman, N., and Caporaso, N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *J. Natl. Cancer Inst. (Bethesda)*, 92: 1151–1158, 2000.
76. Taioli, E., Mari, D., Franceschi, C., Bertolini, S., Marinelli, D., and Garte, S. Distribution of metabolic genes polymorphisms in healthy centenarians. *Biochem. Biophys. Res. Commun.*, 280: 1389–1392, 2001.
77. Smelt, V. A., Mardon, H. J., and Sim, E. Placental expression of arylamine *N*-acetyltransferases: evidence for linkage disequilibrium between *NAT1*10* and *NAT2*4* alleles of the two human arylamine *N*-acetyltransferase loci *NAT1* and *NAT2*. *Pharmacol. Toxicol.*, 83: 149–157, 1999.
78. Ramachandran, S., Lear, J. T., Ramsey, H., Smith, A. G., Bowers, B., Hutchinson, P., Jones, P. W., Fryer, A. A., and Strange, R. C. Presentation with multiple cutaneous basal cell carcinomas: association of glutathione *S*-transferase and cytochrome P450 genotypes with clinical phenotype. *Cancer Epidemiol Biomark. Prev.*, 8: 61–67, 1999.
79. Wikman, H., Thiel, S., Jager, B., Schmezer, P., Spiegelhalder, B., Elder, L., Dienemann, H., Kayser, K., Schulz, V., Drings, P., Bartsch, H., and Risch, A. Relevance of *N*-acetyltransferases 1 and 2 (*NAT1*, *NAT2*) genetic polymorphisms in non-small lung cancer. *Pharmacogenetics*, 11: 157–168, 2001.